A few things I learned at the 6th USENIX Conference on File and Storage Technologies (FAST'08)

Dan Gunter

11 March 2008

50 ways to lose your data

- Media fail, e.g., CD's can go bad in as little as 18 months, formats go out of date, ..
- But the issue is not just losing data
  - finding data: remembering where/how to search
  - remembering you even had it
    - if you don't remember you had it, do you need to find it?
- And not just "local storage". People have stuff scattered all over the net
- More info: Cathy Marshall's home page http://research.microsoft.com/~cathymar/

Spy vs. Spy: TOCTTOU races

- TOCTTOU stands for “time of check to time of use”
- fascinating “filesystem maze” technique that is vaguely reminiscent of DSD symlink practices :-)
- Explains why UNIX access() manpage says:
  - `Access() is a potential security hole and should never be used.`
- More info:
  - http://www.cs.huji.ac.il/~dants/papers/Tocttou08FAST.pdf

Petascale = World of Pain (even if you do get funding)

- PDSI BOF, Garth Gibson reiterated Dan Reed's thesis except from a storage perspective:
  - Petascale system recovery strategy, checkpointing, does not scale because at 0.5M nodes, a disk will fail during every checkpoint
  - Shrinking memory only buys a little time
- Also worth checking out: John Shalf's parallel I/O measurements: http://web-dev.nersc.gov/news/presentations/Shan_CUG07.pdf

File storage has gone "OO"

- Object Storage Devices (OSD)
  - higher-level interface than a "block"
  - objects contain data and attributes
    - analogous to a filesystem *inode*
  - wire protocol goes over iSCSI (an embedding of SCSI-3 over TCP/IP)
    - standard is being developed as we speak
- Panasas clustered file system (commercial)
  - Layered on OSD; files striped over 2+ objects
  - http://www.usenix.org/event/fast08/tech/full_papers/welch/welch.pdf

That’s all folks

- I have the full proceedings on CD if anyone wants it